Lecture 2: The Ergodic Theorem

Review of Measure Theory: A probability space $(\Omega, \mathcal{F}, \mu)$ is made of

- A set Ω ("phase / sample / config space")
- A collection $\mathcal{F}$ of measurable sets, with the σ-algebra axioms: (i) $\emptyset, \Omega \in \mathcal{F}$; (ii) closed under complements; (iii) closed under countable unions and intersections
- A probability measure $\mu : \mathcal{F} \to [0,1]$ s.t. $\mu(\Omega) = 1$ and with the σ-additivity property:

$$\left.\begin{array}{l} E_1, E_2, E_3, \ldots \in \mathcal{F} \\ E_i \text{ pairwise disjoint} \end{array}\right\} \Longrightarrow \mu\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mu(E_i)$$

Terminology: A property $P(\omega)$ of $\omega \in \Omega$ holds μ-almost everywhere ("μ-a.e") if the set $A = \{\omega \in \Omega : P(\omega) \text{ holds}\}$ is measurable and $\mu(A) = \mu(\Omega) = 1$.

Equivalently, $A^c = \Omega \setminus A = \{\omega \in \Omega : P(\omega) \text{ doesn't hold}\}$ is measurable, of measure zero. But it need not be empty

Example: Suppose μ is the measure on $[0,1]$ $\mu = \delta_{1/2}$, i.e.

$$\mu(E) = \begin{cases} 1 & 1/2 \in E \\ 0 & 1/2 \notin E \end{cases}.$$

Then $x = 1/2$ μ-a.e. Indeed $\mu\{x \in [0,1] : x \neq 1/2\} = 0$

Probability Preserving Transformations

Def$^{\underline{n}}$. A <u>probability preserving transformation</u> T on a prob space $(\Omega, \mathcal{F}, \mu)$ is a map $T: \Omega \to \Omega$ which is

(a) <u>measurable</u>: For all $E \in \mathcal{F}$, $T^{-1}(E) := \{\omega \in \Omega : T(\omega) \in E\} \in \mathcal{F}$

(b) <u>measure preserving</u>: For every bounded measurable $f: \Omega \to \mathbb{R}$,

$$\int_\Omega f(T\omega)\, d\mu = \int_\Omega f\, d\mu$$

(In particular, $\mu(E) = \mu(T^{-1}E)$, because $1_E \circ T = 1_{T^{-1}E}$.)

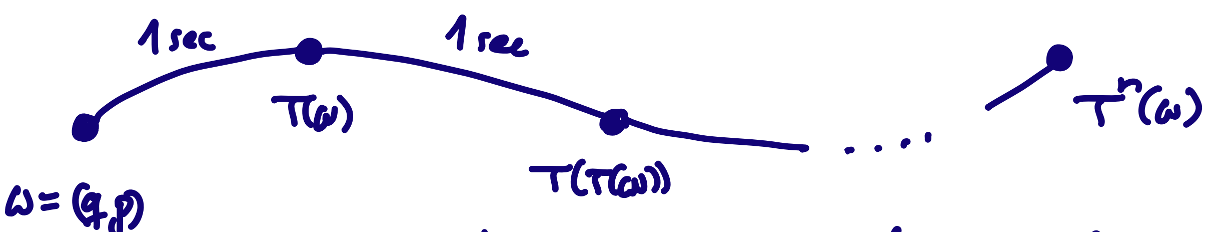
<u>The Principal Example</u>: Consider the law of motion

$$(*)\begin{cases} q_i = -\dfrac{\partial H}{\partial p_i}, \quad p_i = \dfrac{\partial H}{\partial q_i} \qquad (i = 1, \dots, N) \\ H(q_1, \dots, q_N, p_1, \dots, p_N) = \sum_{i=1}^N |p_i|^2 + V(q_1, \dots, q_N) \end{cases}$$

Define $T(q, p) = (q(1), p(1))$ where $(q(t), p(t))$ solves $(*)$ with initial condition $q(0) = q$, $p(0) = p$

<u>Observation</u>: If $T^n = T \circ \dots \circ T$ then $T^n(q, p) = (q(n), p(n))$

This follows from the <u>uniqueness of solution to ODE's</u>



The concatenation of 1 sec forward sol$^{\underline{n}}$s is an n-sec forward solution

<u>Construction</u>: Fix some "large" energy level H_0.

- $\Omega = \{$initial condition$\} = \{(q_1, \dots, q_N; p_1, \dots, p_N) \mid H \le H_0\}$
- $\mathcal{F}$ = the Borel σ-algebra, the smallest σ-algebra which contain all boxes

$$[a_1, b_1) \times \cdots \times [a_{6N}, b_{6N})$$

- μ = the (unique) measure on $\mathcal{F}$ s.t.

$$\mu([a_1, b_1] \times \cdots \times [a_{6N}, b_{6N}]) = \prod_{i=1}^{6N} (b_i - a_i) \Big/ Z$$

$(Z = \text{volume of } \{(q,p) : H(q,p) \le H_0\})$.

Mathematics: "Lebesgue measure" $d\mu$

Physics: "Liouville measure", "$dp\,dq$"

<u>Liouville Theorem</u>: $T: \Omega \to \Omega$ is a probability preserving map on $(\Omega, \mathcal{F}, \mu)$.

<u>Remarks</u>: In this example

- $T^n(\omega) = (T \circ \cdots \circ T)(\omega)$ = state of the system at time n when the initial state was ω
- If $f(q,p)$ is a <u>conserved quantity</u> (e.g. $H(q,p)$) then $f \circ T = f$. Such functions are called <u>invariant</u>.

The Ergodic Theorem

<u>Def</u>: A probability preserving transformation is called <u>ergodic</u>, if every <u>a.e. invariant</u> measurable function is <u>a.e. constant</u>:

(a) <u>a.e. invariant</u>: $f(T\omega) = f(\omega)$ μ-a.e.
(i.e. $\mu\{\omega \in \Omega : f(T\omega) \neq f(\omega)\} = 0$).

(b) <u>a.e. constant</u>: for some constant $c \in \mathbb{R}$, $f = c$ μ-a.e.
(i.e. $\mu\{\omega \in \Omega : f(\omega) \neq c\} = 0$).

<u>Morally speaking</u>: Invariant functions = conserved quantities

<u>Crucial difference</u>: Potentially, there are many more μ-a.e. measurable invariant functions, than <u>globally defined</u> <u>continuous</u> conserved quantities

> To check ergodicity it is <u>not</u> enough to show that all continuous globally defined invariant functions are constant.

This is why checking ergodicity is so difficult.

The Ergodic Thm (Birkhoff, '31): Suppose T is a probability preserving map on a probability space $(\Omega, \mathcal{F}, \mu)$, and let $f: \Omega \to \mathbb{R}$ be a measurable function s.t. $\int |f| d\mu < \infty$.

(1) The limit $\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} f(T^k(\omega))$ exists μ-a.e.

(2) If T is ergodic $\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} f(T^k(\omega)) = \int_\Omega f d\mu$ μ-a.e.

Corollary: Suppose T is ergodic, then for every $E \in \mathcal{F}$, for μ-a.e. $\omega \in \Omega$

$$\frac{1}{N} \# \{ 1 \le n \le N : T^n(\omega) \in E \} \xrightarrow[N\to\infty]{} \mu(E).$$

Proof. Apply to $f = 1_E = \begin{cases} 1 & \text{on } E \\ 0 & \text{outside } E \end{cases}$.

The Main Problem with the Ergodic Theorem

Let

$$\Omega_\mu(f) = \left\{ \omega \in \Omega : \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(T^n\omega) = \int f d\mu \right\}.$$

The ergodic thm says that $\mu[\Omega_\mu(f)^c] = 0$, and $\mu[\Omega_\mu(f)] = 1$. But in general $\Omega_\mu(f) \ne \Omega$, and for a given $\omega \in \Omega$, we don't know how to decide if $\omega \in \Omega_\mu(f)$ or not.

<u>Fact of Life</u> : It is very common for chaotic maps to have many different ergodic invariant measures.

If μ_1, μ_2 are two such measures, and

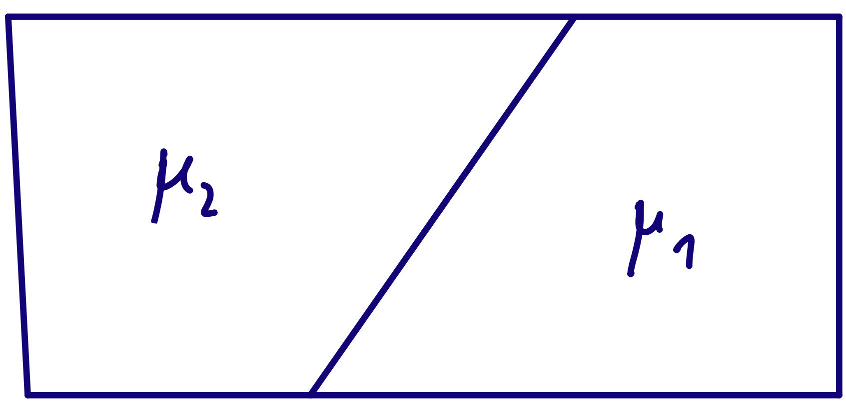
$$\int_\Omega f \, d\mu_1 \neq \int_\Omega f \, d\mu_2$$

(e.g. $f = 1_E$ and $\mu_1(E) \neq \mu_2(E)$) then

$$\mu_i \left[\Omega_{\mu_i}(f) \right] = 1 \qquad \text{by the ergodic thm}$$

but $\mu_1(\Omega_{\mu_2}(f)) = \mu_2(\Omega_{\mu_1}(f)) = 0$, because

$$\underbrace{\Omega_{\mu_1}(f)}_{\lim = \int f d\mu_1} \cap \underbrace{\Omega_{\mu_2}(f)}_{\lim = \int f d\mu_2} = \emptyset$$



For a given ω, how to know if ω is μ_1-typical or μ_2-typical, or neither?

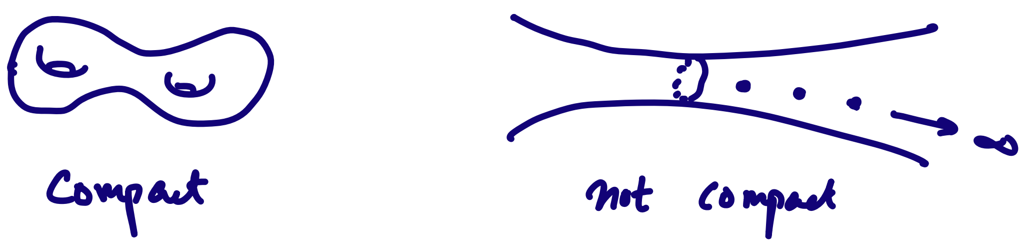
Unique Ergodicity

Setup: A metric space is a set X with a distance function $d(x,y)$ s.t.

- $d(x,x)=0$
- $d(x,y)=d(y,x)$
- $d(x,y)+d(y,z)\geq d(x,z)$ (triangle inequality)

In this case we say that $x_n \to y$ if $d(x_n,y)\to 0$.

A metric space is called compact if every sequence has a convergent subsequence.



Thm. Let T be a continuous map on a compact metric space Ω. The following are equivalent:

(1) T has exactly one invariant probability measure μ.

(2) For every continuous $f:\Omega\to\mathbb{R}$, for every $\omega\in\Omega$,

$$\frac{1}{N}\sum_{n=1}^{N} f(T^n(\omega)) \xrightarrow[N\to\infty]{} \int_\Omega f\,d\mu$$

In this case we call T uniquely ergodic.

Example: Irrational rotations $T: S^1\to S^1$,

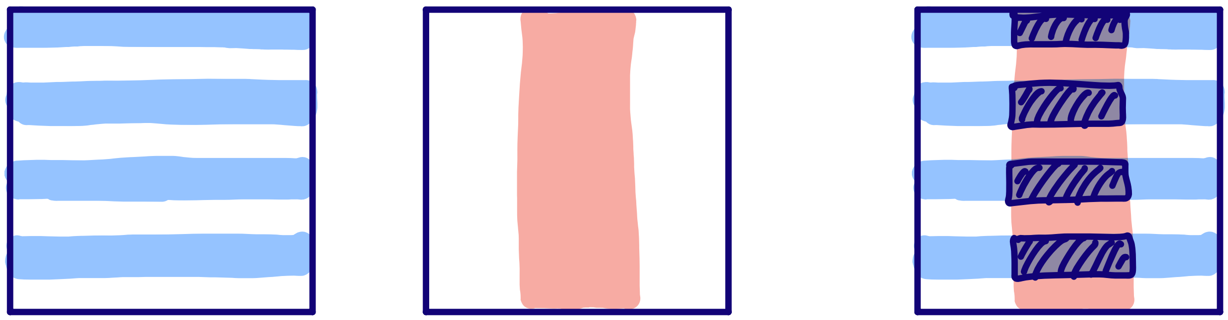
$T(e^{i\theta}) = e^{i(\theta+\alpha)}$, provided $\alpha\notin 2\pi\mathbb{Q}$.

$$\boxed{\text{Most chaotic maps are not uniquely ergodic}}$$

Ergodicity and Mixing

<u>Def</u>. Two measurable events A, B are called <u>independent</u> if $\mu(A \cap B) = \mu(A)\mu(B)$

(because then $\mu(A|B) = \mu(A)$, $\mu(B|A) = \mu(B)$).

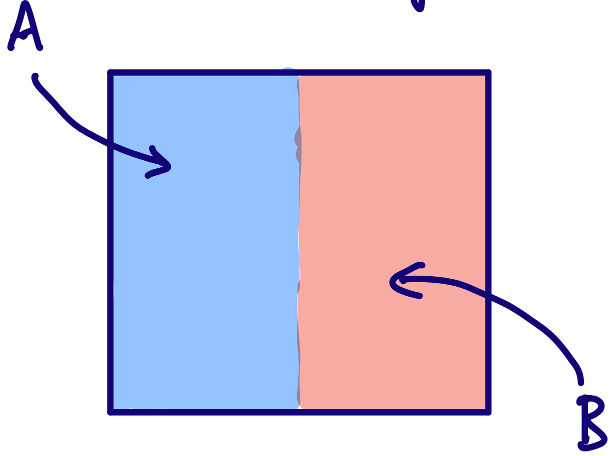


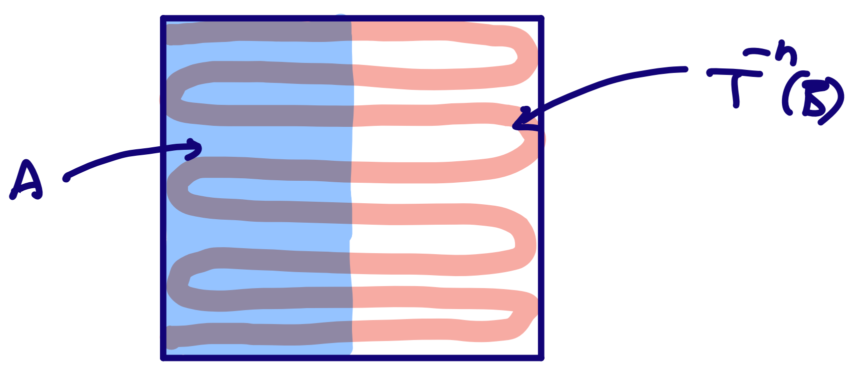
$\mu(A) = 1/2$ $\quad\quad$ $\mu(B) = 1/3$ $\quad\quad$ $\mu(A \cap B) = 1/6$

<u>Def</u>. A probability preserving map T on a prob. space $(\Omega, \mathcal{F}, \mu)$ is called <u>mixing</u>, if for all $A, B \in \mathcal{F}$,

$$\mu(A \cap T^{-n}B) - \mu(A)\mu(B) \xrightarrow[n \to \infty]{} 0.$$

("The event $\underbrace{[T^n(\omega) \in B]}_{\text{future}}$ is asymp. independent of $\underbrace{[\omega \in A]}_{\text{present}}$")





Ergodicity is slightly weaker than mixing:

Thm. A prob preserving map T is ergodic, if and only if for every A, B measurable

$$\frac{1}{N}\sum_{n=1}^{N}\left(\mu(A\cap T^{-n}B)-\mu(A)\mu(B)\right)\xrightarrow[N\to\infty]{}0.$$

What happens if we replace $(\cdots)$ by $|\cdots|$?

Def$\underline{^n}$. A probability preserving map T is called weakly mixing if for all A, B measurable,

$$\frac{1}{N}\sum_{n=1}^{N}\left|\mu(A\cap T^{-n}B)-\mu(A)\mu(B)\right|\xrightarrow[N\to\infty]{}0.$$

Thm (von Neumann): The following are equivalent

(1) T is weak mixing

(2) Every μ-a.e. eigenfunction of T is μ-a.e. constant
(A μ-a.e. eigenfunction: $f:\Omega\to\mathbb{C}$ measurable s.t. $f\circ T=\lambda f$ μ-a.e. with λ constant).

(3) For every A, B measurable, there's $\mathcal{N}_{AB}\subseteq\mathbb{N}$ of density zero, s.t.

$$\mu(A\cap T^{-n}B)-\mu(A)\mu(B)\xrightarrow[\mathcal{N}_{AB}\not\ni n\to\infty]{}0$$

ergodicity	no inv function	$\frac{1}{N}\sum_{1}^{N}\left(\mu(A\cap T^{-n}B)-\mu(A)\mu(B)\right)\to 0$
weak mixing	no eigenfunction	$\frac{1}{N}\sum_{1}^{N}\left|\mu(A\cap T^{-n}B)-\mu(A)\mu(B)\right|\to 0$
mixing	—	$\mu(A\cap T^{-n}B)-\mu(A)\mu(B)\to 0$